Determinantal Correlations for Classical Projection Processes

Peter J. Forrester[†] and Taro Nagao[*]

[†]*Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia*

[*] *Graduate School of Mathematics, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan*

Abstract

Recent applications in queuing theory and statistical mechanics have isolated the process formed by the eigenvalues of successive minors of the GUE. Analogous eigenvalue processes, formed in general from the eigenvalues of nested sequences of matrices resulting from random corank 1 projections of classical random matrix ensembles, are identified for the LUE and JUE. The correlations for all these processes can be computed in a unified way. The resulting expressions can then be analyzed in various scaling limits. At the soft edge, with the rank of the minors differing by an amount proportional to $N^{2/3}$, the scaled correlations coincide with those known from the soft edge scaling of the Dyson Brownian motion model.

1 Introduction

Since the pioneering days of Wigner, Gaudin, Mehta and Dyson (see Porter [31] for a collection of papers from this period), random matrix theory has shown itself to be perhaps the richest source of exact solutions for correlation and distribution functions of all statistical mechanical models. Many of the discoveries of this type have their motivation in new applications of random matrix theory, unknown in the pioneering days. A case in point is the recent work of Johansson and Nordenstam [22, 27], who compute the exact form of the correlation functions for the coupled eigenvalue sequences obtained from the principal minors of Gaussian unitary ensemble (GUE) matrices.

The motivation for studying this GUE minor process begins with a work of Baryshnikov [1]. Some years earlier Glynn and Whitt [15] had studied the problem of computing the distribution of exit times from a queueing system, in the limit the number of queues tends to infinity but the number of jobs remains finite. It was proved that for general i.i.d. service times the scaled distributions D_k of the exit time of the kth customer from the final queue could be written as

$$D_k = \sup_X \sum_{i=0}^{k-1} \Big(B_i(t_{i+1}) - B_i(t_i) \Big).$$

Here each B_i denotes an independent standard Brownian motion, while the condition X is that

$$0 = t_0 < t_1 < \cdots < t_k = 1.$$

By studying the particular case of exponential waiting times Baryshnikov was able to show that $\{D_k\}_{k=1,2,\dots}$ could alternatively be specified as the joint distribution of $\{\mu_k\}$, where μ_k is the largest eigenvalue of the kth principal minor of an infinite GUE matrix X with probability density function (PDF) proportional to $\exp(-X^2/2)$. Johansson and Nordenstam [22] give other occurrences in statistical mechanics of essentially the same process relating to the eigenvalues of minors of GUE matrices (referred to as the GUE minor process). These are in the specification of certain point processes relating to the boundary region (neighbourhood of the frozen zone) of random domino tilings of the Aztec diamond [20], and to random lozenge-tilings of a hexagon [21]. In the equivalent language of stepped surfaces, Okounkov and Reshetikhin [29] make similar observations. The recent work of Borodin, Ferrari and Sasamoto [5] encounters this process in the context of studying the dynamics of the asymmetric exclusion process.

The broader setting of the measures encountered in the studies of the above statistical mechanical models relates to the Robinson-Schensted-Knuth (RSK) correspondence (see Section 2.1) below. Consideration of the structures inherent therein [2, 14] identifies natural extensions of the GUE minor process. One such class of extensions replaces the Gaussian weight in the latter by the Laguerre or Jacobi weights, which can all be realized by a sequence of projections onto random complex hyperplanes, giving rise to so called classical projection processes (the Gaussian, Laguerre and Jacobi weights are all classical from the viewpoint of the theory of orthogonal polynomials). Our main point in the present paper is that the correlations for the classical projection processes can be computed exactly in a unified way. The essential ingredient here is the Rodrigues formula for classical orthogonal polynomials. Moreover, known asymptotic formulas for the latter allow for the evaluation of scaling limits of the correlations.

We begin in Section 2 by recalling how the RSK correspondence relates to a certain statistical mechanical model of last passage times. In Section 3 the occurrence of special cases of the joint PDF in some random matrix setting is noted. The correlations for the classical projection process are calculated in Section 4, and their scaling limits analyzed in Section 5.

2 A joint probability density associated with RSK

2.1 The case of general parameters

The Robinson-Schensted-Knuth (RSK) correspondence gives a bijection between $n_1 \times n_2$ non-negative integer matrices $[x_{i,j}]$ (rows countered from the bottom) with entry (ij) weighted $(a_i b_j)^{x_{i,j}}$, and pairs of weighted semi-standard tableaux (weights $\{a_i\}$, $\{b_j\}$) of shape $\mu = (\mu_1, \dots, \mu_n)$). Results from [18, 2] associate a probabilistic model to the RSK correspondence. Identify with each lattice site (ij) a random non-negative integer variable $x_{i,j}$ chosen from the geometric distribution with parameter $a_i b_j$, so that

$$\Pr(x_{i,j} = k) = (1 - a_i b_j)(a_i b_j)^k. \tag{2.1}$$

For given non-negative parameters $\{a_i\}$, $\{b_j\}$, and given $n_1, n_2 \in \mathbb{Z}^+$, a probabilistic quantity of interest is the sequence of last passage times

$$L^{(l)}(n_1, n_2) = \max \sum_{(\mathrm{rd}^*)^l} x_{i,j}, \qquad l = 1, \dots, \max(n_1, n_2). \tag{2.2}$$

Here $(\mathrm{rd}^*)^l$ denotes the set of l disjoint (no common lattice points) rd* lattice paths, which in turn are defined as either a single point, or points connected by segments formed out of arbitrary positive integer multiples of steps to the right and steps up in the rectangle $1 \le i \le n_1$, $1 \le j \le n_2$.

A crucial feature of the RSK correspondence is that the length of the first row of the semi-standard tableaux pair is equal to (2.2) in the case $l = 1$, and thus $\mu_1 = L^{(1)}(n_1, n_2)$. More generally, all row lengths are determined by (2.2) according to [17]

$$\mu_l = L^{(l)}(n_1, n_2) - L^{(l-1)}(n_1, n_2) \tag{2.3}$$

with $L^{(0)}(n_1, n_2) := 0$. Thus the distribution of the last passage times is fully determined by the distribution of $\{\mu_l\}$.

A well known fact relating to the RSK correspondence is that the probability an $n_1 \times n_2$ non-negative matrix with elements chosen according to (2.1) corresponds to a pair of semi-standard tableaux with shape μ, one of content n_1, the other of content n_2, is given by [24]

$$P(\mu) = \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} (1 - a_i b_j) s_\mu(a_1, \dots, a_{n_1}) s_\mu(b_1, \dots, b_{n_2}), \tag{2.4}$$

where s_μ denotes the Schur polynomial. A lesser known fact is that for $n_1 > n_2$ the joint probability that an $n_1 \times (n_2 + 1)$ non-negative integer matrix with elements chosen according to (2.1) corresponds to a pair of semi-standard tableaux with shape μ, content n_1 and $n_2 + 1$, and the bottom left sub-matrix corresponds to a pair of semi-standard tableaux with shape κ, content n_1 and n_2 is [14]

$$\prod_{i=1}^{n_1} \prod_{j=1}^{n_2+1} (1 - a_i b_j) s_\mu(a_1, \dots, a_{n_1}) s_\kappa(b_1, \dots, b_{n_2}) b_{n_2+1}^{\sum_{j=1}^{n_2}(\mu_j - \kappa_j) + \mu_{n_2+1}} \chi(\mu > \kappa) \tag{2.5}$$

where, with $\chi(A) = 1$ if A is true, $\chi(A) = 0$ otherwise,

$$\chi(\mu > \kappa) := \chi(\mu_1 \ge \kappa_1 \ge \mu_2 \ge \kappa_2 \ge \cdots \ge \mu_{n_1} \ge \kappa_{n_1} \ge 0). \tag{2.6}$$

It follows from (2.4) and (2.5) that given the pair of semi-standard tableaux corresponding to an $n_1 \times n_2$ matrix ($n_1 > n_2$) has shape κ, the probability of the pair of semi-standard tableaux corresponding to the $n_1 \times (n_2 + 1)$ matrix, obtained by adding an extra row to the existing matrix, having shape μ is

$$P(\mu, \kappa) := \chi(\mu > \kappa) \prod_{i=1}^{n_1} (1 - a_i b_{n_2+1}) \frac{s_\mu(a_1, \dots, a_{n_1})}{s_\kappa(a_1, \dots, a_{n_1})} b_{n_2+1}^{\sum_{j=1}^{n_2}(\mu_j - \kappa_j) + \mu_{n_2+1}}. \tag{2.7}$$

Let us now seek the joint probability that with $n_1 \geq n_2 + p$, an $n_1 \times (n_2 + p)$ non-negative integer matrix with elements chosen according to (2.1) is such that the principal $n_1 \times (n_2 + s)$ sub-blocks ($s = 0, 1, \dots, p$) correspond to pairs of semi-standard tableaux with shape $\mu^{(s)}$. This is computed from (2.4) and (2.7) according to

$$\begin{aligned} P(\mu^{(0)}) \prod_{s=0}^{p-1} P(\mu^{(s+1)}, \mu^{(s)}) &= \prod_{i=1}^{n_1} \prod_{j=1}^{n_2+p} (1 - a_i b_j) s_{\mu^{(p)}}(a_1, \dots, a_{n_1}) s_{\mu^{(0)}}(b_1, \dots, b_{n_2}) \\ &\quad \times \prod_{s=1}^{p} b_{n_2+s}^{\sum_{j=1}^{n_2+s-1}(\mu_j^{(s)} - \mu_j^{(s-1)}) + \mu_{n_2+s}^{(s)}} \chi(\mu^{(s)} > \mu^{(s-1)}). \end{aligned} \tag{2.8}$$

2.2 Specializing the parameters

In [14] the joint probability (2.5) is specialized to the case of a geometrical progression of parameters

$$\begin{aligned} (a_1, \dots, a_{n_1}) &= (z, zt, zt^2, \dots, zt^{n_1-1}) \\ (b_1, \dots, b_{n_2}) &= (z, zt, zt^2, \dots, zt^{n_2-1}). \end{aligned} \tag{2.9}$$

This is a preliminary step for taking the so called Jacobi limit, in which a joint probability of a type known from random matrix theory is obtained. In this subsection the parameters will be specialized according to (2.9) for the more general joint probability (2.8).

Now for (2.8) to be non-zero we require $\ell(\mu^{(p)}) \leq n_2 + p$. Under this circumstance, we deduce from [14, Eq. (2.26)] with $\kappa \mapsto \mu^{(p)}$, $n_2 \mapsto n_1$, $n_1 \mapsto n_2 + p$, $r_j \mapsto h_j^{(p)} := \mu_j^{(p)} + n_2 + p - j$ that

$$\begin{aligned} s_{\mu^{(p)}}(1, t, \dots, t^{n_1-1}) &= t^{-\sum_{j=1}^{n_1-n_2-p} j(j-1)} t^{-(n_2+p)\sum_{j=1}^{n_1-n_2-p} j} \frac{t^{-\sum_{j=1}^{n_1}(j-1)(n_2+p-j)}}{\prod_{l=1}^{n_1-1}(t;t)_l} \\ &\quad \times \prod_{i=1}^{n_1-n_2-p-1} (t;t)_i \prod_{i=1}^{n_2+p} \frac{(t;t)_{h_i^{(p)}+n_1-n_2-p}}{(t;t)_{h_i^{(p)}}} \prod_{i<j}^{n_2+p} (t^{h_j^{(p)}} - t^{h_i^{(p)}}). \end{aligned} \tag{2.10}$$

This makes explicit the first Schur polynomial factor in (2.8). In regards to the second Schur polynomial factor, making use of [14, Eq. (2.18)] with $n \mapsto n_2$, $n^* \mapsto n_2$, $\lambda \mapsto \mu^{(0)}$, $h_j \mapsto h_j^{(0)} := \mu_j^{(0)} + n_2 - j$ gives

$$s_{\mu^{(0)}}(1, t, \dots, t^{n_2-1}) = \frac{t^{-\sum_{j=1}^{n_2}(j-1)(n_2-j)}}{\prod_{l=1}^{n_2-1}(t;t)_l} \prod_{i<j}^{n_2} (t^{h_j^{(0)}} - t^{h_i^{(0)}}). \tag{2.11}$$

Use of these results in (2.8) allows the following results to be deduced.

Proposition 1. *Let $n_1 \geq n_2 + p$. On each site of the $n_1 \times (n_2 + p)$ square lattice specify a random non-negative integer $x_{i,j}$ according to the specification*

$$\begin{aligned} \Pr(x_{i,j} = k) &= (1 - z^2 t^{i+j-2})(z^2 t^{i+j-2})^k, \quad j \leq n_2 \\ \Pr(x_{i,n_2+s} = k) &= (1 - \alpha_s z t^{i-1})(\alpha_s z t^{i-1})^k \quad (s = 1, \ldots, p). \end{aligned}$$

Introduce the notations $h_i^{(p)} := \mu_i^{(p)} + n_2 + p - i$ and

$$\tilde{\chi}(h^{(p)}, h^{(p-1)}) := \chi(h_1^{(p)} \geq h_1^{(p-1)} > h_2^{(p)} \geq h_2^{(p-1)} > \cdots > h_{n_2+p-1}^{(p)} \geq h_{n_2+p-1}^{(p-1)} > h_{n_2+p}^{(p)}).$$

In this setting, the joint probability that the matrix $[x_{i,j}]_{n_1 \times (n_2+p)}$ is such that the sub-matrices $[x_{i,j}]_{n_1 \times (n_2+s)}$ $(s = 0, \ldots, p)$ correspond, under RSK, to pairs of tableaux of shape $\mu^{(s)}$ has the explicit form

$$\begin{aligned} &K_{n_1,n_2,p}(\{\alpha_s\}, z, t) z^{\sum_{j=1}^{n_2+p} h_j^{(p)} + \sum_{j=1}^{n_2} h_j^{(0)}} \prod_{s=1}^{p} \alpha_s^{\sum_{j=1}^{n_2+s} h_j^{(s)} - \sum_{j=1}^{n_2+s-1} h_j^{(s-1)}} \tilde{\chi}(h^{(s)}, h^{(s-1)}) \\ &\times \prod_{i=1}^{n_2+p} \frac{(t;t)_{h_i^{(p)}+n_1-n_2-p}}{(t;t)_{h_i^{(p)}}} \prod_{i<j}^{n_2+p} (t^{h_j^{(p)}} - t^{h_i^{(p)}}) \prod_{i<j}^{n_2} (t^{h_j^{(0)}} - t^{h_i^{(0)}}) \end{aligned} \tag{2.12}$$

with

$$\begin{aligned} K_{n_1,n_2,p}(\{\alpha_j\}, z, t) &= z^{-2\sum_{j=1}^{n_2}(n_2+p-j)} z^{-\sum_{s=1}^{p}(p-s)} \Big(\prod_{s=1}^{p} \alpha_s^{s-p} \Big) t^{-\sum_{j=1}^{n_1-n_2-p} j(j-1)} \\ &\times t^{-(n_2+p)\sum_{j=1}^{n_1-n_2-p} j} \frac{t^{-\sum_{j=1}^{n_2}(j-1)(n_2-j)}}{\prod_{l=1}^{n_2-1}(t;t)_l} \frac{t^{-\sum_{j=1}^{n_1}(j-1)(n_2+p-j)}}{\prod_{l=1}^{n_1-1}(t;t)_l} \\ &\times \prod_{l=1}^{n_1-n_2-p-1} (t;t)_l \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} (1 - z^2 t^{i+j-2}) \prod_{s=1}^{p} \prod_{i=1}^{n_1} (1 - \alpha_s z t^{i-1}). \end{aligned}$$

2.3 The Jacobi limit

The Jacobi limit of the setting of Proposition 1 corresponds to each site of the $n_1 \times (n_2+p)$ square lattice specifying a non-negative continuous exponential random variable with site dependent variance $j \leq n_2$ given by

$$\begin{aligned} \Pr(x_{i,j} \in [y, y+dy]) &= (i + j - 2 + 2a) e^{-y(i+j-2+2a)} dy, \qquad j \leq n_2 \\ \Pr(x_{i,n_2+s} \in [y, y+dy]) &= (i - 1 + a + a_s) e^{-y(i-1+a+a_s)} dy, \qquad (s = 1, \ldots, p). \end{aligned} \tag{2.13}$$

This can be obtained from (2.12) by setting

$$t = e^{-1/L}, \quad z = e^{-a/L}, \quad \alpha_s = e^{-a_s/L}, \quad h_j^{(s)}/L = x_j^{(s)} \tag{2.14}$$

and taking the limit $L \to \infty$. The joint probability (2.12), scaled by multiplying by $L^{(1+p)(n_2+p/2)}$, has the following limiting form.

Corollary 1. *The PDF obtained by the limiting procedure (2.14) applied to (2.12) is equal to*

$$\begin{aligned} &\tilde{K}_{n_1,n_2,p}(\{a_j\}, a) e^{-a(\sum_{j=1}^{n_2+p} x_j^{(p)} + \sum_{j=1}^{n_2} x_j^{(0)})} \prod_{s=1}^{p} e^{-a_s(\sum_{j=1}^{n_2+s} x_j^{(s)} - \sum_{j=1}^{n_2+s-1} x_j^{(s-1)})} \chi(x^{(s)} > x^{(s-1)}) \\ &\times \prod_{i=1}^{n_2+p} (1 - e^{-x_i^{(p)}})^{n_1-n_2-p} \prod_{1 \leq i<j \leq n_2+p} (e^{-x_j^{(p)}} - e^{-x_i^{(p)}}) \prod_{1 \leq i<j \leq n_2} (e^{-x_j^{(0)}} - e^{-x_i^{(0)}}) \end{aligned} \tag{2.15}$$

where

$$\tilde{K}_{n_1,n_2,p}(\{a_j\},a) = \frac{\prod_{l=1}^{n_1-n_2-p-1} l!}{(\prod_{l=1}^{n_1-1} l!)(\prod_{l=1}^{n_2-1} l!)} \prod_{s=1}^{p} \frac{\Gamma(a_s+a+n_1)}{\Gamma(a_s+a)} \prod_{i=1}^{n_1} \frac{\Gamma(2a+i+n_2-1)}{\Gamma(2a+i-1)}$$

and

$$\chi(x^{(s)} > x^{(s-1)}) := \chi(x_1^{(s)} > x_1^{(s-1)} > \cdots > x_{n_2+s-1}^{(s)} > x_{n_2+s-1}^{(s-1)} > x_{n_2+s}^{(s)} > 0). \tag{2.16}$$

Changing variables $e^{-x_j^{(s)}} = y_j^{(s)}$ the PDF (2.15) reads

$$\tilde{K}_{n_1,n_2,p}(\{a_j\},a) \prod_{i=1}^{n_2+p} (y_i^{(p)})^{a-1}(1-y_i^{(p)})^{n_1-n_2-p} \prod_{j=1}^{n_2} (y_j^{(0)})^a \prod_{s=1}^{p} \Big(\chi(y^{(s)} < y^{(s-1)}) \\ \times \frac{\prod_{j=1}^{n_2+s} (y_j^{(s)})^{a_s}}{\prod_{j=1}^{n_2+s-1} (y_j^{(s-1)})^{a_s+1}} \Big) \prod_{1 \le i < j \le n_2+p} (y_j^{(p)} - y_i^{(p)}) \prod_{1 \le i < j \le n_2} (y_j^{(0)} - y_i^{(0)}) \tag{2.17}$$

where

$$\chi(y^{(s)} < y^{(s-1)}) := \chi(0 < y_1^{(s)} < y_1^{(s-1)} < \cdots < y_{n_2+s-1}^{(s-1)} < y_{n_2+s}^{(s)} < 1). \tag{2.18}$$

In the special case $a_s = a - j$ $(s = 1, \ldots, p)$ this simplifies to the functional form

$$\frac{1}{C} \prod_{l=1}^{n_2+p} w(y_l^{(p)}) \prod_{1 \le i < j \le n_2+p} (y_j^{(p)} - y_i^{(p)}) \prod_{1 \le i < j \le n_2} (y_j^{(0)} - y_i^{(0)}) \prod_{s=1}^{p} \chi(y^{(s)} < y^{(s-1)}), \tag{2.19}$$

with C the normalization (C will be used generally below for this purpose and so its explicit value may vary from equation to equation) and $w(y) = y^\alpha (1-y)^\beta$ for certain α and β. The latter is the Jacobi weight, which in a functional form involving Vandermonde products is typical of a PDF arising in random matrix theory. Indeed, this functional form for each of the Gaussian, Laguerre and Jacobi weights can be obtained as eigenvalue PDFs.

Before turning to such random matrix interpretations, it should be pointed out that in the setting of Proposition 1, choosing each site of the $n_1 \times (n_2+p)$ square lattice according to continuous exponential random variables

$$\Pr(x_{i,j} \in [y, y+dy]) = e^{-y} dy \tag{2.20}$$

leads to (2.19) with the particular Laguerre weight $w(y) = y^{n_1-(n_2+p)} e^{-y}$. Further, rescaling the variables $y_j^{(p)} \mapsto n_1(1 + y_j^{(p)} \sqrt{2/n_1})$ therein, and taking $n_1 \to \infty$ gives (2.19) back with the Gaussian weight $w(y) = e^{-y^2}$. In the case $p = 1$ both of these facts are explicitly demonstrated in [14, Props. 4&5]; the extension to general p involves straightforward limiting procedures applied to (2.17).

3 Joint eigenvalue PDF for some nested sequences of random matrices

3.1 Gaussian unitary ensemble

By definition, matrices M_N from the $N \times N$ GUE satisfy the recurrence

$$M_{N+1} = \begin{bmatrix} M_N & \vec{w} \\ \vec{w}^\dagger & a \end{bmatrix} \tag{3.1}$$

where $a \sim \mathrm{N}[0, 1/\sqrt{2}]$ and each component w_j of $\vec{w}$ has distribution $w_j \sim \mathrm{N}[0, 1/2] + i\mathrm{N}[0, 1/2]$. Further, with U_N the unitary matrix which diagonalizes M_N, and thus $M_N = U_N D_N U_N^\dagger$, where D_N is the diagonal matrix of the eigenvalues of M_N, one has

$$\begin{bmatrix} U_N & \vec{0} \\ \vec{0}^T & 1 \end{bmatrix} \begin{bmatrix} M_N & \vec{w} \\ \vec{w}^\dagger & a \end{bmatrix} \begin{bmatrix} U_N & \vec{0} \\ \vec{0}^T & 1 \end{bmatrix}^\dagger \sim \begin{bmatrix} D_N & \vec{w} \\ \vec{w}^\dagger & a \end{bmatrix}. \tag{3.2}$$

This bordered form is the key to studying the joint distribution of the eigenvalues of the sequence of GUE matrices $\{M_j\}_{j=1,2,\dots}$, or equivalently that of the sequence of principal minors of a single infinite GUE matrix [1, 14, 11].

In particular, it follows from (3.2) that the characteristic polynomials $p_N(\lambda), p_{N+1}(\lambda)$ for M_N, M_{N+1} are related by

$$\frac{p_{N+1}(\lambda)}{p_N(\lambda)} = \lambda - a - \sum_{i=1}^{N} \frac{|w_i|^2}{\lambda - \lambda_i^{(N)}}$$

where $\{\lambda_i^{(N)}\}$ denotes the eigenvalues of M_N assumed ordered

$$\lambda_1^{(N)} < \lambda_2^{(N)} < \cdots < \lambda_N^{(N)}.$$

With $\{\lambda_i^{(N)}\}$ regarded as given the PDF for the zeros of this random rational function, and thus the PDF for the distribution of the eigenvalues $\{\lambda_i^{(N+1)}\}$ of M_{N+1}, can be computed to be equal to

$$e^{-\sum_{j=1}^{N+1}(\lambda_j^{(N+1)})^2} \frac{\prod_{j<k}^{N+1}(\lambda_j^{(N+1)} - \lambda_k^{(N+1)})}{\prod_{j<k}^{N}(\lambda_j^{(N)} - \lambda_k^{(N)})} \chi(\lambda^{(N+1)} < \lambda^{(N)})$$

where $\chi(\lambda^{(N+1)} < \lambda^{(N)})$ is specified as in (2.18) except that the first and last inequalities are replaced by $-\infty <$ and $< \infty$ respectively. From this result, together with the fact that the eigenvalue PDF for $N \times N$ GUE matrices is proportional to

$$\prod_{l=1}^{N} e^{-(\lambda_l^{(N)})^2} \prod_{1 \le j < k \le N} (\lambda_k^{(N)} - \lambda_j^{(N)})^2,$$

it follows that the joint eigenvalue PDF for the sequence of GUE matrices $\{M_n\}_{n=N,\dots,N+p}$ is given by (2.19) with $n_2 = N$ and $w(y) = e^{-y^2}$, and the definition (2.16) of $\chi(y^{(s)} < y^{(s-1)})$ modified appropriately.

3.2 Laguerre unitary ensemble

Matrices $A_{(n)}$ from the $N \times N$ LUE with parameter $a = n - N$ are constructed from $n \times N$ rectangular complex Gaussian matrices $X_{(n)}$ with entries $\mathrm{N}[0, 1/\sqrt{2}] + i\mathrm{N}[0, 1/\sqrt{2}]$ according to $A_{(n)} = X_{(n)}^\dagger X_{(n)}$. Such matrices satisfy the recurrence

$$A_{(n+1)} = A_{(n)} + \vec{x}\vec{x}^\dagger, \qquad A_{(0)} = [0]_{N \times N} \tag{3.3}$$

where $\vec{x}$ is an $N \times 1$ column vector of complex Gaussians. Our interest is in the case $n \le N$ of $A_{(n)}$, for which there are $N - n$ zero eigenvalues. Then, after making use too of the invariance of $\vec{x}\vec{x}^\dagger$ by a unitary similarity transformation, the recursion (3.3) can be written in the equivalent form

$$A_{(n+1)} = \mathrm{diag}(a_1, \dots, a_n, \underbrace{0, \dots, 0}_{N-n}) + \vec{x}\vec{x}^\dagger \tag{3.4}$$

where $\{a_i\}_{i=1,\dots,n}$ are the non-zero eigenvalues of $A_{(n)}$, assumed ordered so that $0 < a_1 < \cdots < a_n$. From (3.4) it follows that the corresponding characteristic polynomials are such that

$$\frac{\det(\lambda 1_N - A_{(n+1)})}{\det(\lambda 1_N - A_{(n)})} = 1 - \sum_{j=1}^{n} \frac{|x_j|^2}{\lambda - a_j} - \frac{\sum_{j=n+1}^{N} |x_j|^2}{\lambda}. \tag{3.5}$$

The conditional PDF for the zeros of this random rational function, and thus the conditional PDF for the eigenvalues of $A_{(n+1)}$, can be computed exactly as [14, 11]

$$\prod_{i=1}^{n+1} \lambda_i^{N-(n+1)} \prod_{j=1}^{n} \frac{1}{a_j^{N-n}} e^{-\sum_{j=1}^{n+1} \lambda_j + \sum_{j=1}^{n} a_j} \frac{\prod_{i<j}^{n+1} (\lambda_i - \lambda_j)}{\prod_{i<j}^{n} (a_i - a_j)} \chi(\lambda < a)$$

Recalling that the non-zero eigenvalue PDF for the matrices $A_{(n)}$ is proportional to

$$\prod_{j=1}^{n} a_j^{N-n} e^{-a_j} \prod_{i<k}^{n} (a_i - a_k)^2$$

it follows that the joint eigenvalue PDF for the sequence of LUE matrices $\{A_{(r)}\}_{r=n,\dots,n+p}$ with $n + p \le N$ is given by (2.19) with $n_2 = n$, $w(y) = y^{N-(n+p)} e^{-y}$.

3.3 Corank 1 random projections

Let A_n be an $n \times n$ matrix with eigenvalues $a_1 < a_2 < \cdots < a_n$, and let $\vec{x}$ be an $n \times 1$ random complex Gaussian normalized column vector. The matrix

$$M_n := \Pi_n A_n \Pi_n, \qquad \Pi_n := 1_n - \vec{x}\vec{x}^\dagger \tag{3.6}$$

then represents a corank 1 random projection of A_n. We know from [14] that in general M_n has a single zero eigenvalue, while the non-zero eigenvalues $\{\lambda_j\}_{j=1,\dots,n-1}$ have the conditional PDF

$$(n-1)! \frac{\prod_{i<j}^{n-1} (\lambda_j - \lambda_k)}{\prod_{i<j}^{n} (a_j - a_k)} \chi(a < \lambda). \tag{3.7}$$

Introduce a nested sequence of matrices $\{A_i\}_{i=1,\dots,n}$ by setting A_{n-1} equal to the diagonal matrix formed from the non-zero eigenvalues of M_n, computing M_{n-1} according to (3.6), setting A_{n-2} equal to the diagonal matrix formed from the the non-zero eigenvalues of M_{n-1}, and repeating. With the eigenvalues of A_n having PDF proportional to

$$\prod_{l=1}^{n} w(a_l) \prod_{1 \le j < k \le n} (a_k - a_j)^2 \tag{3.8}$$

it follows immediately from (3.7) that the joint PDF for $\{A_i\}_{i=n-p,\dots,n}$ is given by (2.19) with $n_2 + p = n$.

We remark (see e.g. [8]) that the Laguerre and Jacobi cases can be realized by the matrix structure $A_n = X_n^\dagger X_n$ for certain matrices (recall the discussion of the previous subsection for the Laguerre case). It follows that the above construction is then equivalent to applying a sequence of corank 1 projections directly to the matrix X_n.

4 Correlation for the classical projection process

4.1 Approach via a general formula

In studying correlations associated with (2.19), the most general case occurs when $n_2 = 1$. After then changing notation by setting $p = N-1$, $y_l^{(p)} \mapsto x_{p+2-l}^{(p+1)}$, (2.19) reads

$$\frac{1}{C}\prod_{l=1}^{N} w(x_l^{(N)}) \prod_{1\le j<k\le N}(x_j^{(N)} - x_k^{(N)}) \prod_{s=1}^{N-1} \chi(x^{(s+1)} > x^{(s)}). \tag{4.1}$$

Here $\chi(x^{(s+1)} > x^{(s)})$ is defined as

$$\chi(x^{(s+1)} > x^{(s)}) = \chi(x_1^{(s+1)} > x_1^{(s)} > \cdots > x_s^{(s+1)} > x_s^{(s)} > x_{s+1}^{(s+1)})$$

(cf. (2.16)) and $w(x)$ involves a factor $\chi_{0<x<1}$. Hereafter we consider more general cases in which $w(x)$ does not always involve such a factor (see (4.7) below). Of course the product of differences can be written in terms of the Vandermonde determinant, giving

$$\prod_{l=1}^{N-1} w(x_l^{(N)}) \prod_{1\le j<k\le N}(x_j^{(N)} - x_k^{(N)}) \propto \det[w(x_k^{(N)})p_{N-j}(x_k^{(N)})]_{j,k=1,\dots,N},$$

with $\{p_j(x)\}_{j=0,\dots,N-1}$ a set of arbitrary polynomials, $p_j(x)$ of degree j. Furthermore, we know from [13, Lemma 1] that

$$\chi(x^{(s+1)} > x^{(s)}) = \det[\chi(x_j^{(s+1)} > x_k^{(s)})]_{j,k=1,\dots,s+1} \tag{4.2}$$

where $x_{s+1}^{(s)} := -\infty$. Consequently (4.1) can be written in the form

$$\frac{1}{C}\prod_{s=1}^{N-1} \det[\phi(x_j^{(s)}, x_k^{(s+1)})]_{j,k=1,\dots,s+1} \det[\Psi_{N-j}^{N}(x_k^{(N)})]_{j,k=1,\dots,N} \tag{4.3}$$

with

$$\phi(x,y) := \chi_{y>x}, \qquad \Psi_j^N(x) := w(x)p_j(x). \tag{4.4}$$

The general structure (4.3) is precisely that for which the correlations have been determined in the recent work [4, Lemma 3.4]. To apply this result, with $(a*b)(x,y) := \int_{-\infty}^{\infty} a(x,z)b(z,y)\,dz$, it is necessary to compute the quantities

$$\phi^{(n_1,n_2)}(x,y) := (\underbrace{\phi * \cdots * \phi}_{n_2-n_1 \text{ times}})(x,y), \qquad n_1 < n_2$$

(for $n_1 \ge n_2$, $\phi^{(n_1,n_2)}(x,y) := 0$), and

$$\Psi_{n-j}^{n}(x) := (\phi^{(n,N)} * \Psi_{N-j}^{N})(x) \quad (1 \le n < N,\ j = 1,\dots,N).$$

Use of (4.4) shows

$$\phi^{(n_1,n_2)}(x,y) = \frac{1}{(n_2-n_1-1)!}\chi_{y>x}(y-x)^{n_2-n_1-1} \tag{4.5}$$

(with the convention that $1/(-p)! = 0$ for $p \in \mathbb{Z}^+$, this vanishes for $n_1 \geq n_2$) and

$$\Psi^n_{n-j}(x) = \frac{1}{(N-n-1)!} \int_x^\infty w(y) p_{N-j}(y)(y-x)^{N-n-1}\, dy. \tag{4.6}$$

To proceed further, we choose $w(y)$ to be one of the classical weight functions

$$w(y) = \begin{cases} e^{-y^2}, & \text{Gaussian} \\ y^a e^{-y} \chi_{y>0}, & \text{Laguerre} \\ y^a (1-y)^b \chi_{0<y<1}, & \text{Jacobi.} \end{cases} \tag{4.7}$$

We further choose $p_j(y)$ to be proportional to the corresponding orthogonal polynomials, as specified by their Rodrigues formulas

$$p_j(y) = \frac{1}{e_j w(y)} \frac{d^j}{dy^j} \Big(w(y)(Q(y))^j \Big) = \begin{cases} H_j(y), & \text{Gaussian} \\ L^{(a)}_j(y), & \text{Laguerre} \\ P^{(a,b)}_j(1-2y), & \text{Jacobi} \end{cases} \tag{4.8}$$

with the quantities e_j and $Q(y)$ defined in the various cases by the pairs

$$(e_j, Q(y)) = \begin{cases} ((-1)^j, 1), & \text{Gaussian} \\ (j!, y), & \text{Laguerre} \\ (2^j j!, y(1-y)), & \text{Jacobi.} \end{cases} \tag{4.9}$$

Substituting (4.8) in (4.6) and integrating by parts shows that for $j \geq 0$ ($n \neq N$)

$$\Psi^n_j(x) = (-1)^{N-n} \frac{e_j}{e_{N-n+j}} \begin{cases} w(x) H_j(x), & \text{Gaussian} \\ w(x)|_{a \mapsto a+N-n} L^{(a+N-n)}_j(x), & \text{Laguerre} \\ w(x)|_{\substack{a \mapsto a+N-n \\ b \mapsto b+N-n}} P^{(a+N-n,b+N-n)}_j(1-2x), & \text{Jacobi,} \end{cases} \tag{4.10}$$

while for $j < 0$

$$\Psi^n_j(x) = \frac{(-1)^{N-n+j}}{e_{N-n+j}} \frac{1}{(-j-1)!} \int_x^\infty (y-x)^{-j-1} w(y)(Q(y))^{N-n+j}\, dy. \tag{4.11}$$

As further required by [4, Lemma 3.4], one introduces the polynomials $\{\Phi^n_j(x)\}_{j=0,\dots,n-1}$, $n = 1, \dots, N-1$ by the orthogonality requirement

$$\int_{-\infty}^\infty \Phi^n_j(x) \Psi^n_k(x)\, dx = \delta_{j,k}.$$

From (4.10) we see that

$$\Phi^n_j(x) = (-1)^{N-n} \frac{e_{N-n+j}}{e_j} \begin{cases} \frac{1}{\mathcal{N}_j} H_j(x), & \text{Gaussian} \\ \frac{1}{\mathcal{N}_j|_{a \mapsto a+N-n}} L^{(a+N-n)}_j(x), & \text{Laguerre} \\ \frac{1}{\mathcal{N}_j|_{\substack{a \mapsto a+N-n \\ b \mapsto b+N-n}}} P^{(a+N-n,b+N-n)}_j(1-2x), & \text{Jacobi,} \end{cases} \tag{4.12}$$

where

$$\mathcal{N}_j = \begin{cases} 2^j j! \sqrt{\pi}, & \text{Gaussian} \\ \frac{\Gamma(j+a+1)}{\Gamma(j+1)}, & \text{Laguerre} \\ \frac{\Gamma(j+a+1)\Gamma(j+b+1)}{j!(2j+a+b+1)\Gamma(j+a+b+1)}, & \text{Jacobi} \end{cases} \tag{4.13}$$

is the normalization associated with each polynomial respectively.

A crucial feature exhibited by (4.12) is that $\Phi_0^n(x)$ is a constant. Now, Assumption (A) of [4, Lemma 3.4] requires that $\Phi_0^n(x) \propto \phi(x_{n+1}^{(n)}, x)$. Recalling from below (4.2) that $x_{n+1}^{(n)} := -\infty$, we see from the first definition in (4.4) that indeed $\phi(x_{n+1}^{(n)}, x)$ is similarly a constant. With Assumption (A) satisfied, [4, Eq. (3.25)] gives that the correlation between eigenvalues of species s_j at positions y_j $(j = 1, \dots, r)$ has the determinant form

$$\rho(\{(s_j, y_j)\}_{j=1,\dots,r}) = \det[K(s_j, y_j; s_k, y_k)]_{j,k=1,\dots,r}, \tag{4.14}$$

with the kernel K given in terms of the quantities $\phi^{(n_1,n_2)}(x, y)$, $\Psi_j^n(x)$, $\Phi_j^n(x)$ specified above according to

$$K(s_j, y_j; s_k, y_k) = -\phi^{(s_j,s_k)}(y_j, y_k) + \sum_{l=1}^{s_k} \Psi_{s_j-l}^{s_j}(y_j)\Phi_{s_k-l}^{s_k}(y_k). \tag{4.15}$$

These findings can be summarized in the following statement.

Proposition 2. *Consider the joint PDF (4.1), with $w(y)$ one of the three classical weights (4.7). Specify $\phi^{(n_1,n_2)}(x, y)$ by (4.5); $e_j, Q(y)$ by (4.9); $\Psi_j^n(x)$ by (4.10), (4.11); $\Phi_j^n(x)$ by (4.12) and $\mathcal{N}_j$ by (4.13). In terms of these quantities, the general r-point correlation is given by (4.14) with kernel (4.15).*

4.2 Direct approach

Here the method of [26] will be used to reclaim Proposition 2. The starting point is (4.3), rewritten in the form

$$\frac{1}{C}\prod_{s=2}^{N} \det\begin{bmatrix} 1_{(N-s)\times(N-s)} & 0_{(N-s)\times s} \\ 0_{s\times(N-s)} & \begin{bmatrix} [\phi(x_j^{(s-1)}, x_k^{(s)}) - \kappa_s(x_j^{(s-1)})]_{\substack{j=1,\dots,s-1\\ k=1,\dots,s}} \\ [1]_{k=1,\dots,s} \end{bmatrix} \end{bmatrix}$$
$$\times \det[\psi_{j-1}^N(x_k^{(N)})]_{j,k=1,\dots,N} \tag{4.16}$$

so that all determinants are of the same dimension. The auxiliary function $\kappa_s(x)$ is arbitrary, the value of the determinant being independent of $\kappa_s(x)$.

In the notation for $w(y)$, $p_j(y)$ and $\mathcal{N}_j$ as defined by (4.7), (4.8) and (4.13), introduce the superscript $(N-n)$ to indicate that $a \mapsto a+n$ (Laguerre case), $a \mapsto a+n$, $b \mapsto b+n$ (Jacobi case). In the Gaussian case the superscript has no effect. With this meaning understood, expanding $\phi(x, y) = \chi_{y>x}$ in terms of $\{p_j^{(s)}(y)\}_{j=0,1,\dots}$ gives

$$\phi(x, y) = \sum_{k=0}^{\infty} \frac{p_k^{(s)}(y)}{\mathcal{N}_k^{(s)}} \int_x^{\infty} w^{(s)}(t)p_k^{(s)}(t)\, dt. \tag{4.17}$$

Separating off the $k = 0$ term, making use of the Rodrigues formula (4.8) and integrating by parts reduces this to

$$\phi(x, y) = \frac{1}{\mathcal{N}_0^{(s)}} \int_x^{\infty} w^{(s)}(t)\, dt - w^{(s-1)}(x) \sum_{k=0}^{\infty} \frac{e_k}{e_{k+1}\mathcal{N}_{k+1}^{(s)}} p_k^{(s-1)}(x)p_{k+1}^{(s)}(y). \tag{4.18}$$

Recalling from (4.16) that $\kappa_s(x)$ is to be subtracted from $\phi(x, y)$, the formula (4.18) suggests choosing

$$\kappa_s(x) = \frac{1}{\mathcal{N}_0^{(s)}} \int_x^{\infty} w^{(s)}(t)\, dt. \tag{4.19}$$

Making this choice, and with the notation

$$\eta_k^{(s)}(x) = \left(\frac{w^{(s)}(x)}{\mathcal{N}_k^{(s)}} \right)^{1/2} p_k^{(s)}(x) \tag{4.20}$$

(note that $\{\eta_k^{(s)}(x)\}_{k=0,1,\dots}$ is a set of orthonormal functions) we obtain for (4.16) the expression

$$\frac{1}{C} \prod_{s=2}^{N} \det \begin{bmatrix} 1_{(N-s)\times(N-s)} & 0_{(N-s)\times s} \\ 0_{s\times(N-s)} & \begin{bmatrix} [1]_{k=1,\dots,s} \\ [\tilde{\phi}_s(x_j^{(s-1)}, x_k^{(s)})]_{\substack{j=1,\dots,s-1 \\ k=1,\dots,s}} \end{bmatrix} \end{bmatrix}$$
$$\times \det[\eta_{j-1}^{(N)}(x_k^{(N)})]_{j,k=1,\dots,N} \tag{4.21}$$

where, with

$$\gamma_j^{(t)} := e_j (\mathcal{N}_j^{(t)})^{1/2}, \tag{4.22}$$

the quantity $\tilde{\phi}_s$ is specified by

$$\tilde{\phi}_s(x,y) := \sum_{k=0}^{\infty} \frac{\gamma_k^{(s-1)}}{\gamma_{k+1}^{(s)}} \eta_k^{(s-1)}(x) \eta_{k+1}^{(s)}(y)$$

and for convenience the final row in the bottom right block of the first determinant has been moved to the first row.

Our next step is to introduce

$$\eta_{j,l}^{(s)} := \begin{cases} \eta_j^{(s)}(x_l^{(s)}), & j \ge 0,\ l \ge 1 \\ \delta_{j,l-1}, & \text{otherwise} \end{cases}$$

and in terms of this to define

$$\begin{aligned} A_{j,l}^{(s,t)} &:= \sum_{k=0}^{N-1} \frac{\gamma_{k+s-N}^{(s)}}{\gamma_{k+t-N}^{(t)}} \eta_{k+s-N,j+s-N}^{(s)} \eta_{k+t-N,l+t-N}^{(t)} \\ G_{j,l}^{(s,t)} &:= \sum_{k=0}^{\infty} \frac{\gamma_{k+s-N}^{(s)}}{\gamma_{k+t-N}^{(t)}} \eta_{k+s-N,j+s-N}^{(s)} \eta_{k+t-N,l+t-N}^{(t)}. \end{aligned}$$

We can write (4.21) in terms of $\{A_{j,l}^{(s,t)}\}$, $\{G_{j,l}^{(s,t)}\}$ so that it reads

$$\frac{1}{C} \det[A_{j,l}^{(N,1)}]_{j,l=1,\dots,N} \prod_{s=2}^{N} \det[G_{j,l}^{(s-1,s)}]_{j,l=1,\dots,N}$$
$$= \frac{1}{C} \det \begin{bmatrix} A^{(N,1)} & A^{(N,2)} & A^{(N,3)} & A^{(N,4)} & \cdots & A^{(N,N)} \\ 0 & -G^{(1,2)} & -G^{(1,3)} & -G^{(1,4)} & \cdots & -G^{(1,N)} \\ 0 & 0 & -G^{(2,3)} & -G^{(2,4)} & \cdots & -G^{(2,N)} \\ 0 & 0 & 0 & -G^{(3,4)} & \cdots & -G^{(3,N)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -G^{(N-1,N)} \end{bmatrix} \tag{4.23}$$

where $A^{(s,t)} := [A^{(s,t)}_{j,l}]_{j,l=1,\dots,N}$, $G^{(s,t)} = [G^{(s,t)}_{j,l}]_{j,l=1,\dots,N}$. With $\alpha^{(s,t)}$ the $N \times N$ matrix such that $\alpha^{(s,t)}A^{(t,u)} = A^{(s,u)}$, multiply row 1 by $\alpha^{(j-1,N)}$ and add to row j ($j = 2,\dots,N$) to rewrite this as

$$\frac{1}{C}\det\begin{bmatrix} A^{(N,1)} & A^{(N,2)} & A^{(N,3)} & A^{(N,4)} & \dots & A^{(N,N)} \\ A^{(1,1)} & B^{(1,2)} & B^{(1,3)} & B^{(1,4)} & \dots & B^{(1,N)} \\ A^{(2,1)} & A^{(2,2)} & B^{(2,3)} & B^{(2,4)} & \dots & B^{(2,N)} \\ A^{(3,1)} & A^{(3,2)} & A^{(3,3)} & B^{(3,4)} & \dots & B^{(3,N)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A^{(N-1,1)} & A^{(N-1,2)} & A^{(N-1,3)} & A^{(N-1,4)} & \dots & B^{(N-1,N)} \end{bmatrix}$$

where $B^{(s,t)} := A^{(s,t)} - G^{(s,t)}$. Moving the first block-row to the final block-row gives the structured formula

$$\frac{1}{C}\det[F^{(s,t)}]_{s,t=1,\dots,N}, \qquad F^{(s,t)} := \begin{cases} A^{(s,t)}, & s \ge t \\ B^{(s,t)}, & s < t. \end{cases} \tag{4.24}$$

Moreover, from the definition of $A^{(s,t)}$ and $G^{(s,t)}$ we observe that for $s \ge t$

$$F^{(s,t)}_{j,l} = \delta_{j,l}, \qquad j \le N-s \text{ or } l \le N-s$$

while for $s < t$

$$F^{(s,t)}_{j,l} = 0, \qquad j \le N-s \text{ or } l \le N-t.$$

This allows the dimension of the block matrix in (4.24) to be reduced, giving for the joint PDF (4.1), p say,

$$p(\vec{x}^{(1)},\dots,\vec{x}^{(N)}) = \frac{1}{C}\det[f^{(s,t)}]_{s,t=1,\dots,N} \tag{4.25}$$

where $\vec{x}^{(j)} = (x^{(j)}_1,\dots,x^{(j)}_j)$ and $f^{(s,t)}$ is the $s \times t$ matrix with entries

$$f^{(s,t)}_{j,l} = F^{(s,t)}_{j-s+N,l-t+N} = \begin{cases} \sum_{k=1}^{t} \frac{\gamma^{(s)}_{s-k}}{\gamma^{(t)}_{t-k}}\eta^{(s)}_{s-k}(x^{(s)}_j)\eta^{(t)}_{t-k}(x^{(t)}_l), & s \ge t \\ -\sum_{k=-\infty}^{0} \frac{\gamma^{(s)}_{s-k}}{\gamma^{(t)}_{t-k}}\eta^{(s)}_{s-k}(x^{(s)}_j)\eta^{(t)}_{t-k}(x^{(t)}_l), & s < t. \end{cases} \tag{4.26}$$

From the orthonormality of $\{\eta^{(s)}_k(x)\}$ it follows from (4.25) that

$$\int_{-\infty}^{\infty} f^{(s,t)}_{j,l} f^{(t,u)}_{l,m}\, dx^{(t)}_l = \begin{cases} f^{(s,u)}_{j,m}, & s \ge t \ge u \text{ or } s < t < u \\ 0, & \text{otherwise.} \end{cases} \tag{4.27}$$

We seek to use the form (4.25), together with the property (4.27), to compute the correlation between eigenvalues of species s_j at positions y_j ($j = 1,\dots,r$). For this purpose, we group together the eigenvalues of distinct species in the correlation. Thus if the distinct species are $\hat{s}_1,\dots,\hat{s}_{\hat{r}}$, with $\hat{s}_1,\dots,\hat{s}_{\hat{r}} \in \{s_1,\dots,s_r\}$, we write the positions being observed in species $\hat{s}$ as $\vec{x}^{(\hat{s})} := (x^{(\hat{s})}_1,\dots,x^{(\hat{s})}_{n_{\hat{s}}})$ ($1 \le n_{\hat{s}} \le \hat{s}$). The correlation relating to $\{\vec{x}^{(\hat{s}_j)}\}_{j=1,\dots,\hat{r}}$ is specified in terms of the PDF p by

$$\rho(\{\vec{x}^{(\hat{s}_j)}\}_{j=1,\dots,\hat{r}}) = \Big(\prod_{\substack{s=1 \\ s\notin\{\hat{s}_1,\dots,\hat{s}_{\hat{r}}\}}}^{N} \int dx^{(s)}_1 \cdots \int dx^{(s)}_s\Big)\Big(\prod_{a=1}^{\hat{r}} \frac{\hat{s}_a!}{n_{\hat{s}_a}!}\int dx^{\hat{s}_a}_{n_{\hat{s}_a}+1}\cdots\int dx^{(\hat{s}_a)}_{\hat{s}_a}\Big) \times p(\vec{x}^{(1)},\dots,\vec{x}^{(N)}). \tag{4.28}$$

Because of the structure (4.25) and the orthogonality relation (4.27), these integrals can all be computed by performing a Laplace expansion of the determinant (see e.g. [26, 8]) to give

$$\rho(\{\vec{x}^{(\hat{s}_j)}\}_{j=1,\dots,\hat{r}}) = \det\Big[[f_{j,k}^{(\hat{s}_\alpha,\hat{s}_\beta)}]_{\substack{j=1,\dots,n_\alpha \\ k=1,\dots,n_\beta}}\Big]_{\alpha,\beta=1,\dots,\hat{r}}.$$

In the notation of (4.14) this can equivalently be written

$$\rho(\{(s_j,x_j)\}_{j=1,\dots,r}) = \det[f_{j,k}^{(s_j,s_k)}]_{j,k=1,\dots,r} \tag{4.29}$$

(note that the superscript on $x_j^{(s_j)}$ is now redundant, and hence has been omitted) so to complete our task of rederiving (4.14) it is sufficient to show that

$$f_{j,k}^{s_j,s_k} = \frac{a(s_j,x_j)}{a(s_k,x_k)} K(s_j,x_j;s_k,x_k) \tag{4.30}$$

(the corresponding determinant is independent of the function $a(s,x)$).

To verify (4.30) we begin by recalling (4.20) and (4.22) to see from the definition (4.26) that for $s \ge t$

$$f_{j,l}^{(s,t)} = \Big(w^{(s)}(x_j)w^{(t)}(x_l)\Big)^{1/2} \sum_{k=1}^{t} \frac{e_{s-k}}{e_{t-k}} \frac{p_{s-k}^{(s)}(x_j)p_{t-k}^{(t)}(x_l)}{\mathcal{N}_{t-k}^{(t)}} \tag{4.31}$$

and for $s < t$

$$f_{j,l}^{(s,t)} = -\Big(w^{(s)}(x_j)w^{(t)}(x_l)\Big)^{1/2} \sum_{k=-\infty}^{0} \frac{e_{s-k}}{e_{t-k}} \frac{p_{s-k}^{(s)}(x_j)p_{t-k}^{(t)}(x_l)}{\mathcal{N}_{t-k}^{(t)}}. \tag{4.32}$$

On the other hand, it follows from (4.8) and (4.10) that for $j \ge 0$ $(n \ne N)$

$$\psi_j^n(x) = (-1)^{N-n} \frac{e_j}{e_{N-n+j}} w^{(n)}(x) p_j^{(n)}(x),$$

while according to (4.12)

$$\Phi_j^n(x) = (-1)^{N-n} \frac{e_{N-n+j}}{e_j} \frac{p_j^{(n)}(x)}{\mathcal{N}_j^{(n)}}. \tag{4.33}$$

Recalling too that $\phi^{(s,t)}(x,y) = 0$ for $s \ge t$ we then see from the definition (4.15) that for $s \ge t$

$$K(s,x_j;t,x_l) = (-1)^{s-t} w^{(s)}(x_j) \sum_{k=1}^{t} \frac{e_{s-k}}{e_{t-k}} \frac{p_{s-k}^{(s)}(x_j)p_{t-k}^{(t)}(x_l)}{\mathcal{N}_{t-k}^{(t)}}. \tag{4.34}$$

For $s < t$

$$\phi^{(s,t)}(x_j,x_l) = \frac{1}{(t-s-1)!} \chi_{x_l > x_j} (x_l - x_j)^{t-s-1}.$$

Analogous to (4.17), we can expand $\phi^{(s,t)}$ in terms of $\{p_k^{(t)}(y)\}$,

$$\phi^{(s,t)}(x,y) = \frac{1}{(t-s-1)!} \sum_{k=0}^{\infty} \frac{p_k^{(t)}(y)}{\mathcal{N}_k^{(t)}} \int_x^\infty w^{(t)}(u)(u-x)^{t-s-1} p_k^{(t)}(u)\, du.$$

Proceeding now as in the derivation of (4.18) gives

$$\phi^{(s,t)}(x,y) = (-1)^{t-s} w^{(s)}(x) \sum_{k=-\infty}^{s} \frac{e_{s-k}}{\mathcal{N}_{t-k}^{(t)} e_{t-k}} p_{s-k}^{(s)}(x) p_{t-k}^{(t)}(y)$$
$$+ \sum_{k=0}^{t-s-1} \frac{(-1)^k}{(t-s-k-1)!} \frac{p_k^{(t)}(y)}{e_k \mathcal{N}_k^{(t)}} \int_x^\infty w^{(t-k)}(u)(u-x)^{t-s-1-k}\, du. \tag{4.35}$$

But from (4.11), for $j < 0$

$$\Psi_j^n(x) = \frac{(-1)^{N-n+j}}{e_{N-n+j}} \frac{1}{(-j-1)!} \int_x^\infty (u-x)^{-j-1} w^{(n-j)}(u)\, du$$

so after making use too of (4.33) we have

$$\sum_{p=s+1}^{t} \psi_{s-p}^s(x) \Phi_{t-p}^t(y)$$
$$= \sum_{k=0}^{t-s-1} \frac{(-1)^k}{(t-s-k-1)!} \frac{p_k^{(t)}(y)}{e_k \mathcal{N}_k^{(t)}} \int_x^\infty (u-x)^{t-s-1-k} w^{(t-k)}(u)\, dy. \tag{4.36}$$

Thus adding (4.36) to minus (4.35) cancels the final line in the latter. It remains to add to minus (4.35) the quantity

$$\sum_{p=1}^{s} \psi_{s-p}^s(x) \Phi_{t-p}^t(y) = (-1)^{s-t} w^s(x) \sum_{k=1}^{s} \frac{e_{s-k}}{e_{t-k}} \frac{p_{s-k}^s(x) p_{t-k}^t(y)}{\mathcal{N}_{t-p}^{(t)}}. \tag{4.37}$$

This cancels the corresponding terms in the first sum of minus (4.35), giving the result that for $s < t$

$$K(s, x_j; t, x_l) = (-1)^{t-s-1} w^{(s)}(x) \sum_{k=-\infty}^{0} \frac{e_{s-k}}{\mathcal{N}_{t-k}^{(t)} e_{t-k}} p_{s-k}^{(s)}(x) p_{t-k}^{(t)}(y). \tag{4.38}$$

Comparing (4.31) with (4.34), and (4.32) with (4.38), we see that for general s, t

$$f_{j,l}^{(s,t)} = (-1)^{s-t} \left(\frac{w^{(t)}(x_l)}{w^{(s)}(x_j)} \right)^{1/2} K(s, x_j; t, x_l),$$

thus verifying (4.30).

5 Scaling limits

It is well known (see e.g. [9, 8]) that the eigenvalue distributions for the joint PDF (3.8) with the classical weights (4.7) permit three distinct scalings as $n \to \infty$. These correspond to eigenvalues in the bulk of the spectrum, or in the neighbourhood of the spectrum edge. There are two distinct cases of the latter — the soft edge and the hard edge. The hard edge is characterized by the eigenvalue density being strictly zero on one side. This occurs for $x < 0$ in the Laguerre ensemble, and for both $x < 0$ and $x > 1$ in the Jacobi ensemble. In contrast, the neighbourhood of the largest eigenvalue of the Laguerre and Gaussian ensembles is such that the eigenvalue density is to leading order in n zero, but at higher order it is non-zero. This is referred to as a soft edge.

For the projection process (4.1) with classical weights (4.7) we again expect these same three distinct scalings, provided the difference between ranks of the matrices (or equivalently between labels of the species) is fixed. We find at the soft edge the correlations can be interpreted as though the eigenvalues of the different species coincide with the eigenvalues of species (N). Thus they are fully determined by the Airy kernel (see (5.3) below). That the eigenvalues of the different species should coincide at the soft edge is not at all surprising upon consideration of the joint PDF (2.19). Thus one observes that the lowest indexed species repel via a Vandermonde factor, with no restoring potential apart from the ordering constraint. Thus they will tend to cluster at the boundaries, which at the soft edge corresponds to the positions of the species (N).

The situation in the bulk and at the hard edge is more delicate in that the correlations depend on the difference between labels of the species, even though this difference does not scale with N. However the correlations within a given species can be anticipated. With $w(x) = x^a e^{-x}$ in (4.1) the marginal distribution of species $N-c$ is precisely the Laguerre unitary ensemble with $a \mapsto a+c$. Thus the hard edge correlations within species $N-c$ must be the usual Bessel kernel correlations with $a \mapsto a+c$, which is indeed what we find. Because the bulk correlations for the Laguerre and Gaussian unitary ensembles are given by the sine kernel, all bulk correlations within a species will be specified by the usual sine kernel. As to be discussed below, for the bulk scaling it turns out that the full set of correlations are those known from the so called bead process [7].

Soft edge scaling, together with the difference between labels of the species scaling with N is of particular interest for its relevance to the queueing process of Baryshnikov [1], or equivalently a lattice version of the last passage percolation model of Hammersley (see e.g. [10]). To see this, let us revise some aspects of the theory relating to the latter topic.

Thus, with each site (i,j) in the quadrant $\mathbb{Z}^+ \times \mathbb{Z}^+$ an exponential random variable x_{ij} of density $2e^{-2t}$, $t > 0$. Define the stochastic variable

$$l(m,n) = \max \sum_{(1,1)\,\mathrm{u/r}\,(m,n)} x_{ij}, \tag{5.1}$$

where the sum is over all lattice paths in $\mathbb{Z}^+ \times \mathbb{Z}^+$ which start at $(1,1)$ and finish at (m,n) going either one lattice site up (u), or one lattice site to the right (r). It is well known that with $x_n := l(n,n)$ the variables $\{x_1, \dots, x_n\}$ have a joint PDF of the form (3.8) with $w(x) = e^{-x}$ [18]. Hence, as $n \to \infty$ the corresponding distributions permit a soft edge scaling describing the scaled distribution of $l(n,n)$. This scaling is fully described by the correlation function

$$\rho_{(n)}^{\mathrm{soft}}(y_1, \dots, y_n) = \det[K^{\mathrm{soft}}(y_j, y_k)]_{j,k=1,\dots,n} \tag{5.2}$$

where

$$\begin{aligned} K^{\mathrm{soft}}(x,y) &:= \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x-y} \\ &= \int_0^\infty \mathrm{Ai}(x+u)\mathrm{Ai}(y+u)\,du \end{aligned} \tag{5.3}$$

is the so called Airy kernel. It is also known [19] that the sequence of stochastic variables $\{l(n+k, n-k)\}_{k=0,1,\dots}$ permit a scaling to a state specified by the dynamical extension of the correlation (5.2),

$$\rho_{(n)}^{\mathrm{soft}}((\tau_1, y_1), \dots, (\tau_n, y_n)) = \det[K^{\mathrm{soft}}((\tau_j, y_j), (\tau_k, y_k))]_{j,k=1,\dots,n} \tag{5.4}$$

where

$$K^{\rm soft}((\tau_x, x), (\tau_y, y)) = \begin{cases} A^{(1)}_{\tau_y - \tau_x}(x, y), & \tau_y \geq \tau_x \\ A^{(2)}_{\tau_y - \tau_x}(x, y), & \tau_y < \tau_x, \end{cases}$$

$$\begin{aligned} A^{(1)}_\tau(x, y) &:= \int_0^\infty e^{-\tau u} {\rm Ai}(x + u){\rm Ai}(y + u)\, du \\ A^{(2)}_\tau(x, y) &:= -\int_{-\infty}^0 e^{-\tau u} {\rm Ai}(x + u){\rm Ai}(y + u)\, du. \end{aligned} \tag{5.5}$$

This is the so called Airy process $\mathcal{A}_2$, which underlies the distribution of the largest eigenvalue in the scaled limit of the Dyson Brownian motion model of the GUE [25, 12].

The significance of these facts with respect to the present study is that in (2.19) with $n_2 = 0$, $n_1 = p = n$, $w(y) = e^{-y}$, we know from the sentence including (2.20) that the variables $\{y_1^{(n-p)}\}$ (i.e. the largest eigenvalue for each species) coincide with the stochastic variables $\{l(n, n-k)\}_{k=0,1,\dots}$. By analogy with the behaviour of the stochastic variables $\{l(n+k, n-k)\}_{k=0,1,\dots}$, one may anticipate that their distribution is controlled by the Airy process $\mathcal{A}_2$. We will find that with the differences between the ranks (species) scaled to be of order $n^{2/3}$ that this is indeed the case, and that the same effect holds for the soft edge in the Gaussian case.

5.1 Fixed differences between species

Soft edge and bulk scaling

Explicit details will be worked out only in the Gaussian case, as this case is typical; in particular the scaled correlations do not depend on the particular case they originated from as is typical in random matrix theory (a form of universality). In the $N \times N$ GUE the soft edge scaling is obtained by the change of variables

$$x_i = \sqrt{2N} + \frac{X_i}{\sqrt{2}N^{1/6}}. \tag{5.6}$$

This has the effect of moving the origin to the neighbourhood of the largest eigenvalue, and scaling the distances so the inter-eigenvalue spacings in this neighbourhood are of order unity. The bulk scaling is obtained by the change of variables

$$x_i = \frac{\pi X_i}{\sqrt{2N}}, \tag{5.7}$$

which makes the mean particle density in the neighbourhood of the origin unity. With the species differing from N by a constant,

$$s_i = N - c_i, \tag{5.8}$$

we seek the limiting forms of the correlation (4.14) for both the soft edge and bulk scalings.

Proposition 3. *For the soft edge scaling*

$$y_i = \sqrt{2N} + \frac{Y_i}{\sqrt{2}N^{1/6}}, \tag{5.9}$$

and with s_i specified in terms of c_i by (5.8),

$$\frac{1}{\sqrt{2}N^{1/6}} K(s_j, y_j; s_l, y_l) \underset{N\to\infty}{\sim} \frac{a_N(c_j, Y_j)}{a_N(c_l, Y_l)} K^{\rm soft}(Y_j, Y_l), \tag{5.10}$$

where $K^{\rm soft}$ is given by (5.3) and $a_N(c,Y) := e^{-N^{1/3}Y}(2N)^{-c/2}$. Consequently

$$\lim_{N\to\infty}\Big(\frac{1}{\sqrt{2}N^{1/6}}\Big)^r \rho(\{(s_j,y_j)\}_{j=1,\dots,r}) = \det[K^{\rm soft}(Y_j,Y_k)]_{j,k=1,\dots,r}. \tag{5.11}$$

For the bulk scaling

$$y_i = \frac{\pi Y_i}{\sqrt{2N}}, \tag{5.12}$$

and with s_i specified in terms of c_i by (5.8),

$$\frac{\pi}{\sqrt{2N}}K(s_j,y_j;s_l,y_l) \underset{N\to\infty}{\sim} \frac{b_N(c_j)}{b_N(c_l)}B((c_j,Y_j),(c_l,Y_l)), \tag{5.13}$$

where $b_N(c) := (2N)^{-c/2}$ and

$$B((\tau_x,x),(\tau_y,y)) := \left\{ \begin{array}{ll} \displaystyle\int_0^1 s^{\tau_y-\tau_x}\cos(\pi s(x-y)+\pi(\tau_x-\tau_y)/2)\,ds, & \tau_y \ge \tau_x \\ \displaystyle -\int_1^\infty s^{\tau_y-\tau_x}\cos(\pi s(x-y)+\pi(\tau_x-\tau_y)/2)\,ds, & \tau_y < \tau_x \end{array}\right. . \tag{5.14}$$

Consequently

$$\lim_{N\to\infty}\Big(\frac{\pi}{\sqrt{2N}}\Big)^r \rho(\{(s_j,y_j)\}_{j=1,\dots,r}) = \det[B((c_j,Y_j),(c_l,Y_l))]_{j,l=1,\dots,r}. \tag{5.15}$$

Proof. Substituting the Gaussian case of (4.7)–(4.9) and (4.13) in (4.34) shows that for $s_j \ge s_l$ ($c_j \le c_l$)

$$K(s_j,y_j;s_l,y_l) = \frac{e^{-y_j^2}}{\sqrt{\pi}}\sum_{k=1}^{s_l}\frac{1}{2^{s_l-k}(s_l-k)!}H_{s_j-k}(y_j)H_{s_l-k}(y_l). \tag{5.16}$$

Our strategy is to use appropriate expansions of the Hermite polynomials, corresponding to the different scalings, to simplify the summation.

Consider first the soft edge scaling. With x related to X by (5.6) we have the uniform large N expansion [30]

$$e^{-x^2/2}H_N(x) = \pi^{1/4}2^{N/2+1/4}(N!)^{1/2}N^{-1/12}\Big(\mathrm{Ai}(X) + \mathrm{O}(N^{-2/3})\Big\{ \begin{array}{ll} \mathrm{O}(e^{-X}), & X>0 \\ \mathrm{O}(1), & X<0 \end{array}\Big). \tag{5.17}$$

We rewrite this to read

$$\begin{aligned} e^{-x^2/2}H_{N-k}(x) &= \pi^{1/4}2^{(N-k)2+1/4}((N-k)!)^{1/2}N^{-1/12} \\ &\times\Big(\mathrm{Ai}\Big(X+\frac{k}{N^{1/3}}\Big) + \mathrm{O}(N^{-2/3})\Big\{ \begin{array}{ll} \mathrm{O}(e^{-k/N^{1/3}}), & k\ge 0 \\ \mathrm{O}(1), & k<0 \end{array}\Big) \end{aligned} \tag{5.18}$$

and then substitute in (5.16) to obtain

$$\begin{aligned} K(s_j,y_j;s_l,y_l) &\sim e^{-N^{1/3}(Y_j-Y_k)}2^{-(c_j-c_l)/2}2^{1/2}s^{-1/6} \\ &\times\sum_{k=1}^N\Big(\frac{(N-c_j-k)!}{(N-c_l-k)!}\Big)^{1/2}\mathrm{Ai}(Y_j+k/N^{1/3})\mathrm{Ai}(Y_l+k/N^{1/3}). \end{aligned} \tag{5.19}$$

The leading order contribution to the sum in (5.19) comes from terms $\mathrm{o}(k^{1/3})$. Noting that then $(N-c_j-k)!/(N-c_l-k)! \sim N^{c_l-c_j}$ the sum can be recognised as the Riemann sum approximation to the integral form of $K^{\rm soft}$ in (5.3), and (5.10) in the case $c_j \le c_l$ follows.

We turn our attention next to analyzing (5.16) in the bulk scaling limit. For this we use the uniform asymptotic expansion

$$\frac{\Gamma(n/2+1)}{\Gamma(n+1)} e^{-x^2/2} H_n(x) = \cos(\sqrt{2n+1}x - n\pi/2) + \mathrm{O}(n^{-1/2}),$$

and a simple trigonometric identity to deduce that

$$\begin{aligned} K(s_j, y_j; s_l, y_l) \ \sim \ & \frac{1}{2\sqrt{\pi}} \sum_{k=1}^{N-c_l} \frac{1}{2^{N-c_l-k}} \frac{(N-c_j-k)!}{((N-c_j-k)/2)!((N-c_l-k)/2)!} \\ & \times \cos\Big(\pi\sqrt{\frac{N-k}{N}}(Y_j - Y_l) + \frac{\pi}{2}(c_j - c_l)\Big). \end{aligned}$$

Here the main contribution to the sum comes from $(N-k)/N = \mathrm{O}(1)$. Expanding the ratio of factorials in this setting gives

$$\begin{aligned} K(s_j, y_j; s_l, y_l) \ \sim \ & \frac{2^{(c_l-c_j)/2}}{\sqrt{2}\pi} N^{(c_l-c_j-1)/2} \sum_{k=1}^{s_l} \Big(\frac{N-k}{N}\Big)^{(c_l-c_j-1)/2} \\ & \times \cos\Big(\pi\sqrt{\frac{N-k}{N}}(Y_j - Y_l) + \frac{\pi}{2}(c_j - c_l)\Big). \end{aligned} \tag{5.20}$$

Recognizing the sum as the Riemann sum approximation to a definite integral in the variable $(N-k)/N = t$, then changing variables $t = s^2$ in the definite integral gives (5.13) for $c_j \le c_l$.

It remains to study the case $s_j < s_l$ ($c_j > c_l$) for both the hard and soft scalings. For the soft edge scaling it turns out that the form (4.38) is not appropriate. Instead we make use of the form (4.15), which recalling (4.36) and (4.37) can be written

$$\begin{aligned} K(s,x;t,y) = & -\phi^{(s,t)}(x,y) + \sum_{k=0}^{t-s-1} \frac{(-1)^k}{(t-s-k-1)!} \frac{p_k^{(t)}(y)}{e_k \mathcal{N}_k^{(t)}} \int_x^\infty (u-x)^{t-s-1-k} w^{(t-k)}(u)\, du \\ & + (-1)^{s-t} w^{(s)}(x) \sum_{k=1}^{s} \frac{e_{s-k}}{e_{t-k}} \frac{p_{s-k}^{(s)}(x) p_{t-k}^{(t)}(y)}{\mathcal{N}_{t-k}^{(t)}}. \end{aligned} \tag{5.21}$$

Considering this as the sum of three terms, the first two do not contribute in the scaling (5.9), (5.8), and so

$$K(s_j, y_j; s_l, y_l) \ \sim \ (-1)^{s_j - s_l} w^{(s_j)}(y_j) \sum_{k=1}^{s_j} \frac{e_{s_j-k}}{e_{s_l-k}} \frac{p_{s_j-k}^{(s_j)}(y_j) p_{s_l-k}^{(s_l)}(y_l)}{\mathcal{N}_{s_l-k}^{(s_l)}}.$$

This is precisely the expression (4.34), except that the upper terminal is s_j instead of s_l. Recalling the working below (5.16), this detail does not affect the leading asymptotic form, so (5.10) applies for both $c_j \le c_k$ and $c_j > c_k$.

In distinction to the strategy required at the soft edge for $s_j < s_l$ ($c_j > c_l$) , to analyze the bulk scaling in this case the form (4.38) is well suited. As the only difference between (4.38) and (4.34) is in the range of summation, the working leading to (5.20) again applies, so this asymptotic formula remains valid but with k summed from $-\infty$ to 0. Crucially, because $c_j > c_l$ this sum is convergent (it is in relation to this requirement that an analogous approach to the soft edge scaling breaks down), and is furthermore a Riemann sum approximation to the same definite integral as found for the case $c_j \le c_l$, but on $(-\infty, 0]$ instead of $[0,1]$, hence implying the second formula in (5.13). □

5.2 Hard edge scaling

Hard edge scaling is possible for both the Laguerre and Jacobi cases; here the details will be given in the Laguerre case only, as the limiting correlations are the same in both cases. For the $N \times N$ LUE the hard edge scaling results from the change of variables

$$x_i = \frac{X_i}{4N}, \tag{5.22}$$

which makes the inter-eigenvalue spacings in the neighbourhood of the hard edge $x = 0$ of order unity. We seek the limiting correlations with the scaling (5.22) and the species specified by (5.8).

Proposition 4. *For the hard edge scaling (5.22) and with s_i specified in terms of c_i by (5.8)*

$$\frac{1}{4N} K(s_j, x_j; s_l, x_l) \sim \frac{h_N(c_j)}{h_N(c_l)} H(c_j, X_j; c_l, X_l), \tag{5.23}$$

where $h_N(c) := (2N)^{-c}$ and

$$H((\tau_x, x), (\tau_y, y)) := \begin{cases} \frac{1}{4} \int_0^1 s^{(\tau_y - \tau_x)/2} J_{a+\tau_x}((sx)^{1/2}) J_{a+\tau_y}((sy)^{1/2})\, ds, & \tau_y \geq \tau_x \\ -\frac{1}{4} \int_1^\infty s^{(\tau_y - \tau_x)/2} J_{a+\tau_x}((sx)^{1/2}) J_{a+\tau_y}((sy)^{1/2})\, ds, & \tau_y < \tau_x. \end{cases} \tag{5.24}$$

Consequently

$$\lim_{N \to \infty} \Big(\frac{1}{4N}\Big)^r \rho_{(r)}(\{(s_j, x_j)\}_{j=1,\dots,r}) = \det[H(c_j, x_j; c_l, x_l)]_{j,l=1,\dots,N}. \tag{5.25}$$

Proof. Substituting the explicit form of the Laguerre case of the quantities in (4.34) gives

$$K(N - c_j, x_j; N - c_l, x_l) = x_j^{a+c_j} e^{-x_j} \sum_{k=1}^{N-c_l} \frac{\Gamma(N - c_j - k + 1)}{\Gamma(N - k + a + 1)} L_{N-c_j-k}^{(a+c_j)}(x_j) L_{N-c_l-k}^{(a+c_l)}(x_l), \tag{5.26}$$

valid for $c_l \geq c_j$. As x_j, x_l are scaled according to (5.22), it is appropriate to make use of the uniform asymptotic expansion [32]

$$e^{-x/2} x^{a/2} L_n^a(x) = n^{a/2} J_a(2(nx)^{1/2}) + \begin{cases} x^{5/4} \mathrm{O}(n^{a/2-3/4}), & cn^{-1} < x < \omega \\ x^{a/2+2} \mathrm{O}(n^a), & 0 < x < cn^{-1}. \end{cases}$$

Using this, and expanding the ratio of gamma functions with $(N - k)/N = \mathrm{O}(1)$, we deduce

$$\begin{aligned} &K(N - c_j, x_j; N - c_l, x_l) \\ &\quad \sim (2N)^{c_l - c_j} \sum_{k=1}^{N-c_l} \Big(\frac{N-k}{N}\Big)^{(c_l - c_j)/2} J_{a+c_j}\Big(\Big(\frac{N-k}{N} X_j\Big)^{1/2}\Big) J_{a+c_l}\Big(\Big(\frac{N-k}{N} X_l\Big)^{1/2}\Big). \end{aligned} \tag{5.27}$$

This is a Riemann sum, and the result (5.23) in the case $c_l \geq c_j$ follows.

The expression (5.26) is also valid for $c_l < c_j$, provided the summation in now made over $k \in \mathbb{Z}_{\leq 0}$. Following the above working through again gives (5.27), but with the summation over $k \in \mathbb{Z}_{\leq 0}$. Because $c_l < c_j$ the sum is convergent, and its leading form given by the definite integral made explicit in (5.24). ☐

5.3 Soft edge scaling with difference between species $O(N^{2/3})$

As anticipated from the viewpoint of last passage percolation, the soft edge scaling permits well defined correlations with the species separated by $\mathrm{O}(N^{2/3})$. The details can be worked out for both the Gaussian and Laguerre cases, although the limiting correlations correspond to the Airy process $\mathcal{A}_2$ and so are independent of the particular case.

Proposition 5. *In the Gaussian case, scale s_i according to*

$$s_i = N + 2c_i N^{2/3}, \tag{5.28}$$

and scale y_i according to

$$y_i = (2s_i)^{1/2} + \frac{Y_i}{\sqrt{2} s_i^{1/6}}. \tag{5.29}$$

For large N

$$\frac{1}{\sqrt{2}N^{1/6}} K(s_j, y_j; s_l, y_l) \sim \frac{\alpha_N(c_j, Y_j)}{\alpha_N(c_l, Y_l)} \begin{cases} A^{(1)}_{c_j - c_l}(Y_j, Y_l), & c_j \ge c_l \\ A^{(2)}_{c_j - c_l}(Y_j, Y_l), & c_j < c_l \end{cases} \tag{5.30}$$

where $A^{(1)}, A^{(2)}$ are given by (5.5) and $\alpha_N(c, Y) := e^{-N^{1/3}Y}(2N)^{cN^{2/3}} e^{N^{1/3}c^2} e^{-2c^3/3}$. Consequently

$$\lim_{N\to\infty} \Big(\frac{1}{\sqrt{2}N^{1/6}} \Big)^r \rho(\{(s_j, y_j)\}_{j=1,\dots,r}) = \det[K^{\rm soft}((-c_j, X_j), (-c_k, X_k))]_{j,k=1,\dots,r}. \tag{5.31}$$

In the Laguerre case, scale s_i according to

$$s_i = N - \tilde{s}_i, \qquad \tilde{s}_i := 2c_i(2N)^{2/3} \tag{5.32}$$

and scale $y_i^{(s_i)}$ according to

$$y_i^{(s_i)} = 4s_i + 2(a + N - s_i) + 2(2N)^{1/3} Y_i. \tag{5.33}$$

For large N

$$2(2N)^{2/3} K(s_j, y_j; s_l, y_l) \sim \frac{\beta_N(c_j, Y_j)}{\beta_N(c_l, Y_l)} \begin{cases} A^{(1)}_{c_l - c_j}(Y_j, Y_l), & c_l \ge c_j \\ A^{(2)}_{c_l - c_j}(Y_j, Y_l), & c_l < c_j \end{cases} \tag{5.34}$$

with $\beta_N(c, Y) = e^{-(2N)^{1/3}Y} N^{-c(2N)^{2/3}} e^{2(2N)^{1/3}c^2} e^{8c^3/3}$. Consequently

$$\lim_{N\to\infty} \Big(2(2N)^{2/3} \Big)^r \rho(\{(s_j, y_j)\}_{j=1,\dots,r}) = \det[K^{\rm soft}((c_j, Y_j), (c_k, Y_k))]_{j,k=1,\dots,r}. \tag{5.35}$$

Proof. The derivation is very similar in both cases, so we'll be content with presenting the details in the Laguerre case only. Reading off from (5.26) we have

$$K(N - \tilde{s}_j, y_j; N - \tilde{s}_l, y_l) = y_j^{a+\tilde{s}_j} e^{-y_j} \sum_{k=1}^{N-\tilde{s}_l} \frac{\Gamma(N - \tilde{s}_j - k + 1)}{\Gamma(N - k + a + 1)} L^{(a+\tilde{s}_j)}_{N-\tilde{s}_j-k}(y_j) L^{(a+\tilde{s}_l)}_{N-\tilde{s}_l-k}(y_l). \tag{5.36}$$

This formula is valid for $\tilde{s}_j \le \tilde{s}_l$ (for $\tilde{s}_j > \tilde{s}_l$ the RHS is to be modified by multiplying by -1 and changing the summation terminals to $k \in \mathbb{Z}_{\le 0}$; this modification does not change the working below in any essential way, and so will not be considered explicitly).

We seek the asymptotic form of (5.26) upon the scalings (5.33) and (5.32). Adapting the strategy of the proof of Proposition 3, our chief tool is the uniform asymptotic expansion [23]

$$\begin{aligned} x^{a/2}e^{-x/2}L_n^a(x) &= (-1)^n(2n)^{-1/3}\sqrt{(n+a)!/n!} \\ &\times\Big(\mathrm{Ai}(X)+\mathrm{O}(n^{-2/3})\left\{\begin{array}{ll} \mathrm{O}(e^{-X}), & X\geq 0 \\ \mathrm{O}(1), & X<0 \end{array}\right.\Big) \end{aligned} \tag{5.37}$$

where

$$x = 4n+2a+2(2n)^{1/3}X \tag{5.38}$$

(this form allows for $a=\mathrm{o}(n)$; the classical Plancherel-Rotach type formula given in e.g. [32] requires a to be fixed and correspondingly has $\sqrt{(n+a)!/n!}$ replaced by $n^{a/2}$). Use of this formula, rewritten to read

$$\begin{aligned} x^{a/2}e^{-x/2}L_{n-k}^a(x) &= (-1)^{n-k}(2n)^{-1/3}\sqrt{(n-k+a)!/(n-k)!} \\ &\times\Big(\mathrm{Ai}\Big(X+\frac{2k}{(2n)^{1/3}}\Big)+\mathrm{O}(N^{-2/3})\left\{\begin{array}{ll} \mathrm{O}(e^{-k/n^{1/3}}), & k\geq 0 \\ \mathrm{O}(1), & k<0 \end{array}\right.\Big) \end{aligned} \tag{5.39}$$

with $n=N-\tilde{s}_i$ shows that for large N

$$\begin{aligned} K(N-\tilde{s}_j,y_j;N-\tilde{s}_l,y_l) &\sim e^{-(2N)^{1/3}(Y_j-Y_l)} \\ &\times(2N)^{-2/3}\sum_{k=1}^{N-\tilde{s}_l}\Big(\frac{(N-\tilde{s}_j-k)!}{(N-\tilde{s}_l-k)!}\Big)^{1/2}\mathrm{Ai}(Y_j+2k/(2N)^{1/3})\mathrm{Ai}(Y_l+2k/(2N)^{1/3}) \end{aligned}$$

(cf. (5.19)). The leading order contribution to the summation comes from k of order $N^{1/3}$. Using this fact, noting from Stirling's formula that for large s

$$\Big(\frac{(s-k_j)!}{(s-k_l)!}\Big)^{1/2}\sim s^{(k_l-k_j)/2}e^{(k_j^2-k_l^2)/4s}e^{(k_j^3-k_l^3)/12s^2}, \tag{5.40}$$

and using this formula with $s=N$, $k_i=k+\tilde{s}_i$ $(i=j,l)$ the sum is recognised as the Riemann sum approximation to $A^{(1)}$ as defined in (5.5), implying the result (5.34) in the case $c_l\geq c_j$. □

6 Discussion

As pointed out to us by A. Borodin, replacing (5.14) by

$$\left\{\begin{array}{ll} \dfrac{1}{2}\displaystyle\int_{-1}^{1}(is)^{\tau_y-\tau_x}e^{is\pi(x-y)}\,ds, & \tau_y>\tau_x \\ -\dfrac{1}{2}\displaystyle\int_{\mathbb{R}\backslash[-1,1]}(is)^{\tau_y-\tau_x}e^{is\pi(x-y)}\,ds, & \tau_y<\tau_x \end{array}\right. \tag{6.1}$$

leaves the determinant (5.15) unchanged. Further changing scale $x\mapsto x/\pi, y\mapsto y/\pi$ removes the π's from the exponents, and multiplies each integrand by a factor of $1/\pi$ to account for the corresponding scaling of the correlation function. The significance of this form is that it is identical to the $\gamma=0$ case of the correlation kernel for the so-called bead model [7, Thm. 2]. In fact the bead model was already known to be closely related to the GUE minor process [7, Section 4.1]. The form (6.1) can also be obtained as a limit of the incomplete beta kernel of Okounkov and Reshetikhin [28, Section 3.1.7] (write the parameter z as $z=1+ia$, change

variables $w = 1 + ias$, rescale the space variable l by a^{-1} and take $a \to 0$). The recent work [16] obtains the incomplete beta kernel in the context of a study of random lozenge-tilings. Further the kernels of [3, Thm. 4.4] permit degeneracies to (6.1).

Another discussion point is in relation to consistency between the present results, and results from [13]. In [13] the correlations for the p.d.f.

$$\frac{1}{C}\prod_{j=1}^{n} e^{-(x_j+y_j)/2} e^{A(x_y-y_j)/2} \prod_{1\le j<k\le n}(x_j-x_k)(y_j-y_k)\chi_{x_1>y_1>\cdots>x_n>y_n} \tag{6.2}$$

were computed, along with the scaled limits at the soft and hard edges, and in the bulk. The p.d.f. (6.2) with $A=-1$, is identical to the p.d.f. (2.19) with $w(y)=e^{-y}$, $n_2=n$, $p=1$ and $y_{n+1}^{(1)}=0$. Setting $y_{n+1}^{(1)}=0$ would not be expected to alter the soft edge and bulk scaling limits, so it should be that the scaled correlations in [13] contain as special cases the results (5.11) and (5.15) for $|c_j-c_l|=0,1$.

To see that this is indeed the case, we recall from [13] that with

$$A \mapsto \begin{cases} \sqrt{n}\alpha/\pi, & \text{bulk} \\ \alpha/2(2n)^{1/3}, & \text{soft edge,} \end{cases} \tag{6.3}$$

the scaled correlation for k_1 variables of species type x, and k_2 variables of species type y was calculated to equal

$$\begin{aligned} &\rho_{(k_1,k_2)}(X_1,\dots,X_{k_1};Y_1,\dots,Y_{k_2}) \\ &\quad = \det\begin{bmatrix} [K_{\rm oo}^{\rm scaled}(X_j,X_l)]_{j,l=1,\dots,k_1} & [K_{\rm oe}^{\rm scaled}(X_j,Y_l)]_{\substack{j=1,\dots,k_1\\ l=1,\dots,k_2}} \\ [K_{\rm eo}^{\rm scaled}(Y_j,X_l)]_{\substack{j=1,\dots,k_2\\ l=1,\dots,k_1}} & [K_{\rm ee}^{\rm scaled}(Y_j,Y_l)]_{j,l=1,\dots,k_2} \end{bmatrix} \end{aligned} \tag{6.4}$$

where

$$\begin{aligned} K_{\rm ee}^{\rm scaled}(Y,Y') &= K^{\rm scaled}(Y,Y') \\ K_{\rm eo}^{\rm scaled}(Y,X) &= -e^{\alpha(X-Y)}\chi_{X>Y} + e^{\alpha X/2}\int_{-\infty}^{X} e^{-\alpha v/2}K^{\rm scaled}(v,Y)\,dv \\ K_{\rm oe}^{\rm scaled}(X,Y) &= -e^{-\alpha X/2}\frac{\partial}{\partial X}\Big(e^{\alpha X/2}K^{\rm scaled}(X,Y)\Big) \\ K_{\rm oo}^{\rm scaled}(X,X') &= -e^{\alpha(X-X')/2}\frac{\partial}{\partial X}\Big(e^{\alpha X/2}\int_{-\infty}^{X'} e^{-\alpha v/2}K^{\rm scaled}(X,v)\,dv\Big). \end{aligned} \tag{6.5}$$

In the soft edge case $K^{\rm scaled} = K^{\rm soft}$ as specified by (5.3), while in the bulk $K^{\rm scaled} = K^{\rm bulk}$ where

$$K^{\rm bulk}(X,Y) = \frac{\sin\pi(X-Y)}{\pi(X-Y)} = \int_0^1 \cos\pi(X-Y)t\,dt \tag{6.6}$$

We see from (6.3) that $A=-1$ corresponds to $\alpha=0$ in the bulk, and $\alpha\to-\infty$ at the soft edge. We see from (6.5) that for $\alpha\to-\infty$

$$\begin{aligned} &K_{\rm eo}^{\rm soft}(Y,X) \sim -\frac{2}{\alpha}K^{\rm soft}(Y,X), \qquad K_{\rm oe}^{\rm soft}(X,Y) \sim -\frac{\alpha}{2}K^{\rm soft}(X,Y), \\ &K_{\rm oo}^{\rm soft}(X,X') \sim K^{\rm soft}(X,X'). \end{aligned}$$

When substituted in (6.4) the factors $-2/\alpha$, $-\alpha/2$ cancel, and so agreement with (5.11) is found. Further, setting $\alpha = 0$ in (6.5) and recalling (6.6) gives

$$K_{\rm eo}^{\rm bulk}(Y,X)\Big|_{\alpha=0} = -\chi_{X>Y} + \int_{-\infty}^{X} \frac{\sin \pi(v-Y)}{\pi(v-Y)}\, dv,$$
$$K_{\rm oe}^{\rm bulk}(X,Y)\Big|_{\alpha=0} = \pi \int_0^1 t \sin \pi(X-Y)t\, dt, \qquad K_{\rm oo}^{\rm scaled}(X,X')\Big|_{\alpha=0} = \frac{\sin \pi(X-X')}{\pi(X-X')}.$$

A simple calculation shows that the first of these can be rewritten

$$K_{\rm eo}^{\rm bulk}(Y,X)\Big|_{\alpha=0} = -\int_1^{\infty} \frac{\sin \pi v(X-Y)}{\pi v}\, dv.$$

With this we obtain agreement with (5.15) in the case $|c_j - c_l| = 0, 1$, as expected.

Acknowledgements

We thank Eric Nordenstam for some stimulating discussions, and Alexei Borodin for the content of the first paragraph on Section 6. The work of PJF is supported by the Australian Research Council.

Appendix

Since the completion of this work, Borodin and Péché have posted a work [6] on the arXiv which, amongst other results, establishes our Proposition 5. The strategy used is, at a technical level, quite different to that adopted here.

In this appendix we concern ourselves with another aspect of the work [6], relating to a generalization of our (3.3),

$$A_{(n+1)} = A_{(n)} + \vec{x}_{(n)} \vec{x}_{(n)}^{\dagger}, \qquad A_{(0)} = [0]_{p\times p} \tag{A.1}$$

where $\vec{x}_{(n)}$ is a $p \times 1$ column vector of complex Gaussians with entries such that the modulus of the i-th component has distribution $\Gamma[1, 1/(\pi_i + \hat{\pi}_n)]$. As in Section 3.2, the point of interest is in the joint eigenvalue PDF for $\{A_1, \dots, A_p\}$. This is not computed directly, but as in the discussion around (2.20), it is noted that the directed percolation in the $p \times p$ square which each lattice site (i,j) containing an exponential random variable of density $(\pi_i + \hat{\pi}_j)e^{-(\pi_i+\hat{\pi}_j)s}$ has the distribution of the stochastic variable $l(p,p)$ equal to that of the distribution of the largest eigenvalue of $A_{(p)}$. For the percolation problem, the joint distribution of $\{l(j,p)\}_{j=1,\dots,p}$ can be calculated, leading to a joint PDF for the p species of variables $\{x_j^{(s)}\}$, $(s = 1, \dots, p)$ with $j = 1, \dots, s$ proportional to

$$\det[e^{-\pi_i x_j^{(p)}}]_{i,j=1,\dots,p} \prod_{k=1}^{p-1} \det\Big[e^{-\hat{\pi}_{k+1}(x_j^{(k+1)} - x_i^{(k)})} \chi_{x_j^{(k+1)} > x_i^{(k)}} \Big]_{i,j=1,\dots,k+1} e^{-\hat{\pi}_1 x_1^{(1)}}. \tag{A.2}$$

The question is asked as to whether this joint PDF is in fact the joint PDF for the eigenvalues of the matrices $A_{(s)}$, $s = 1, \dots, p$.

In fact the working from [14, Section 5] allows this question to be answered in the affirmative in the limit $\pi_i \to c$, $(i = 1, \dots, p)$. Thus it follows from [14, Corollary 3] that the condition PDF

for the eigenvalues $\{a_j\}_{j=1,\dots,n}$ of $A_{(n)}$ is proportional to

$$\prod_{i=1}^{n+1} \lambda_i^{p-(n+1)} \prod_{j=1}^{n} \frac{1}{a_j^{p-n}} e^{-(c+\hat{\pi}_n)(\sum_{j=1}^{n+1}\lambda_j - \sum_{j=1}^{n} a_j)} \frac{\prod_{i<j}^{n+1}(\lambda_j-\lambda_i)}{\prod_{i<j}^{n}(a_j-a_i)} \chi(\lambda < a). \tag{A.3}$$

Let us now write $\lambda_j \mapsto \lambda_j^{(n+1)}$, $a_j \mapsto \lambda_j^{(n)}$. The sort joint PDF is the product from $n = 1,\dots,p-1$ of the conditional PDFs (A.3), multiplied by the PDF in the case $n=1$, which is proportional to $(x_1^{(1)})^{p-1}e^{-(c+\hat{\pi}_1)x_1^{(1)}}$. This gives (A.2) with the first determinant therein replaced by $\prod_{l=1}^{p} e^{-cx_l} \prod_{i<j}^{p}(x_j^{(p)} - x_i^{(p)})$, thus verifying (A.2) in the case that $\pi_i \to c$, $(i=1,\dots,p)$.

References

[1] Y. Baryshnikov, *GUEs and queues*, Probab. Theory Relat. Fields **119** (2001), 256–274.

[2] J. Baik and E.M. Rains, *Algebraic aspects of increasing subsequences*, Duke Math. J. **109** (2001), 1–65.

[3] A. Borodin, *Periodic Schur process and cylindric partitions*, arXiv:math/0601019.

[4] A. Borodin, P.L. Ferrari, M. Prähoffer, and T. Sasamoto, *Fluctuation properties of the TASEP with periodic initial configuration*, arXiv:math-ph/0608056.

[5] A. Borodin, P.L. Ferrari, and T. Sasamoto, *Transition between Airy*$_1$ *and Airy*$_2$ *processes and the TASEP fluctuations*, arXiv:math-ph/0703023.

[6] A. Borodin and S. Péché, *Airy kernel with two sets of parameters in directed percolation and random matrix theory*, arXiv:0712.1086v1[math-ph].

[7] C. Boutillier, *The bead model and limit behaviours of dimer models*, arXiv:math/0607162v1 [math.PR].

[8] P.J. Forrester, *Log-gases and Random Matrices*, www.ms.unimelb.edu.au/˜matpjf/matpjf.html.

[9] ———, *The spectrum edge of random matrices*, Nucl. Phys. B **403** (1993), 709–728.

[10] ———, *Growth models, random matrices and Painlevé transcendents*, Nonlinearity **16** (2003), R27–R49.

[11] ———, *Beta random matrix ensembles*, To appear in proceedings of the IMS, NUS, 2007.

[12] P.J. Forrester, T. Nagao and G. Honner, *Correlations for the orthogonal-unitary and symplectic-unitary transitions at the hard and soft edges*, Nucl. Phys. B **553** (1999), 601–643.

[13] P.J. Forrester and E.M. Rains, *Correlations for superpositions and decimations of Laguerre and Jacobi orthogonal matrix ensembles with a parameter*, Probab. Theory Relat. Fields **130** (2004), 518–576.

[14] ———, *Interpretations of some parameter dependent generalizations of classical matrix ensembles*, Probab. Theory Relat. Fields **131** (2005), 1–61.

[15] P. Glynn and W. Whitt, *Departures form many queues in series*, Ann. Appl. Probability **1** (1991), 546–572.

[16] V. Gorin, *Non-intersecting paths and Hahn polynomial ensembles*, arXiv:0708.234v1 [math.PR].

[17] C. Greene, *An extension of Schensted's theorem*, Adv. in Math. **14** (1974), 254–265.

[18] K. Johansson, *Shape fluctuations and random matrices*, Commun. Math. Phys. **209** (2000), 437–476.

[19] ———, *Discrete polynuclear growth and determinantal processes*, Commun. Math. Phys. **242** (2003), 277–329.

[20] ———, *The artic circle boundary and the Airy process*, Ann. Probab. **33** (2005), 1–30.

[21] ———, *Non-intersecting, simple, symmetric random walks and the extended Hahn kernel*, Ann. Inst. Fourier **55** (2005), 2129–2145.

[22] K. Johansson and E. Nordenstam, *Eigenvalues of GUE minors*, Elect. J. Probability **11** (2006), 1342–1371.

[23] I.M. Johnstone, *On the distribution of the largest principal component*, Ann. Stat. **29** (2001), 295–327.

[24] D.E. Knuth, *Permutations, matrices and generalized Young tableaux*, Pacific J. Math. **34** (1970), 709–727.

[25] A.M.S. Macêdo, *Universal parametric correlations at the soft edge of the spectrum edge of random-matrix ensembles*, Europhysics Letters **26** (1994), 641–646.

[26] T. Nagao and P.J. Forrester, *Multilevel dynamical correlation functions for Dyson's Brownian motion model of random matrices*, Phys. Lett. A **247** (1998), 42–46.

[27] E.J.G. Nordenstam, *Erratum to Eigenvalues of GUE minors*, Elect. J. Probability **12** (2007), 1048–1051.

[28] A. Okounkov and N. Reshetikhin, *Correlation function of Schur process with applications to local geometry of a random 3-dimensional Young diagram*, J. Amer. Math.Soc. **16** (2003), 581–603.

[29] A. Okounkov and N. Reshetikhin, *The birth of a random matrix*, Moscow Math. J. **6** (2006), 553–566.

[30] F. Olver, *Asymptotics and Special Functions*, Academic Press, London, 1974.

[31] C.E. Porter, *Statistical theories of spectra: fluctuations*, Academic Press, New York, 1965.

[32] G. Szegö, *Orthogonal polynomials*, American Mathematical Society, Providence R.I., 4th edtion, 1975.